HOMEWORK 1 SOLUTIONS

Handed out: Tuesday, Aug. 29, 2017 Due: Wednesday, Sept. 13 midnight

Solution 1. For the matrix $\mathbf{\Sigma}$ to be a covariance matrix, it needs to be symmetric (which obviously is the case) and semi-definite positive, i.e. for every $\boldsymbol{x} \in \mathbb{R}^2, \boldsymbol{x}^T\mathbf{\Sigma}\boldsymbol{x} \geq 0$ or for every (x_1, x_2) the following needs to hold:

$$\sigma^2 x_1^2 + 2\omega\sigma\tau x_1 x_2 + \tau^2 x_2^2 = (\sigma x_1 + \omega\tau x_2)^2 + \tau^2 x_2^2(1-\omega^2) \geq 0$$

A sufficient condition for $\mathbf{\Sigma}$ to be semi-positive definite is that $\det\mathbf{\Sigma} = \sigma^2\tau^2(1-\omega^2) \geq 0$ which is equivalent to $|\omega| \leq 1$. We herein assume that $\sigma \neq 0$ and $\tau \neq 0$. In this case, $\boldsymbol{x}^T\mathbf{\Sigma}\boldsymbol{x} \geq 0$. The matrix $\mathbf{\Sigma}$ is further non-singular if $\det\mathbf{\Sigma} > 0$, which is equivalent to $|\omega| < 1$.

The conditional distribution of x_2 given x_1 equals to the value of the joint density $f(x_1, x_2)$ over the marginal density $f(x_1)$, so let's start with calculating the joint density:

$$\begin{aligned} f(x_1,x_2) =& \frac{1}{2\pi(\det\mathbf{\Sigma})^{1/2}} \exp\left[-\frac{1}{2}\begin{pmatrix} x_1-\mu_1 & x_2-\mu_2 \end{pmatrix} \mathbf{\Sigma}^{-1} \begin{pmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{pmatrix} \right] \\ =& \frac{1}{2\pi\sigma\tau\sqrt{(1-\omega^2)}} \exp\left[-\frac{1}{2(1-\omega^2)} \left(\frac{(x_1-\mu_1)^2}{\sigma^2} + \frac{(x_2-\mu_2)^2}{\tau^2} - 2\omega\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma\tau} \right) \right] \end{aligned}$$

Dividing the joint density by the marginal one $f(x_1) = \frac{1}{\sqrt{2\pi}\sigma}\exp(-(x_1-\mu_1)^2/2\sigma^2)$, we can derive

$$\begin{aligned} f(x_2|x_1) =& f(x_1,x_2)/f(x_1) = \frac{1}{\sqrt{2\pi}\tau\sqrt{(1-\omega^2)}} \\ & \cdot \exp\left[-\frac{1}{2(1-\omega^2)\tau^2}\left((x_1-\mu_1)^2\frac{\tau^2\omega^2}{\sigma^2} + (x_2-\mu_2)^2 - 2\omega\tau\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma} \right) \right] \\ =& \frac{1}{\sqrt{2\pi}\tau\sqrt{(1-\omega^2)}} \exp\left[-\frac{1}{2(1-\omega^2)\tau^2}\big(x_2-\mu_2-\omega\tau/\sigma(x_1-\mu_1)\big)^2 \right] \end{aligned}$$

In conclusion, the conditional distribution of x_2 given x_1 is also a Gaussian distribution with $\mathcal{N}\Big(\mu_2 + \omega\tau/\sigma(x_1-\mu_1), (1-\omega^2)\tau^2\Big)$

p.s. One can also start with the joint distribution, then keep x_1 constant and complete the square on x_2.

Solution 2. The solution procedure goes as follows.

- **Univariate normal distribution**:

  The density is $p(y|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}\exp(-\frac{(y-\mu)^2}{2\sigma^2})$, which can be re-written as

  $$p(y|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{y^2}{2\sigma^2}+\frac{y\mu}{\sigma^2}-\frac{\mu^2}{2\sigma^2}-\log\sigma\right)$$

  It is clear that $h(y) = \frac{1}{\sqrt{2\pi}}$, $\theta = [\theta_1,\theta_2] = [\mu/\sigma^2, -1/2\sigma^2]$, $R(y) = [y, y^2]$, $\Psi(\theta) = \mu^2/2\sigma^2 + \log\sigma = -\theta_1^2/4\theta_2 - \frac{1}{2}\log(-2\theta_2)$.

- **Binomial distribution**:

  The density is $p(y|n,p) = \binom{n}{y}p^y(1-p)^{n-y}$, which can be re-written as

  $$p(y|n,p) = \binom{n}{y}\exp\left(y\log\frac{p}{1-p}+n\log(1-p)\right)$$

  It is clear that $h(y) = \binom{n}{y}$, $\theta = \log(\frac{p}{1-p})$, $R(y) = y$, $\Psi(\theta) = -n\log(1-p) = n\log(1+e^\theta)$.

- **Geometric distribution**:

  The density of the distribution (corresponding to the number of failures before a success) can be written as $p(y|p) = p(1-p)^y$. For such case

  $$p(y|p) = \exp\left(y\log(1-p)+\log p\right)$$

  It is clear that $h(y) = 1$, $\theta = \log(1-p)$, $R(y) = y$, $\Psi(\theta) = -\log p = -\log(1-e^\theta)$.

- **Poisson distribution**:

  The density is $p(y|\lambda) = \frac{\lambda^y\exp(-\lambda y)}{y!}$, which can be re-written as

  $$p(y|\lambda) = \frac{1}{y!}\exp\left(y\log\lambda-\lambda\right)$$

It is clear that $h(y) = \frac{1}{y!}$, $\theta = \log\lambda$, $R(y) = y$, $\Psi(\theta) = \lambda = \exp(\theta)$.

- **Exponential distribution**:

  The density is $p(y|\lambda) = \lambda\exp(-\lambda y)$, which can be re-written as

$$p(y|\lambda) = \lambda\exp(-\lambda y) = \exp\Big(-\lambda y + \log(\lambda)\Big)$$

  and it is clear that $h(y) = 1$, $\theta = \lambda$, $R(y) = -y$, $\Psi(\theta) = -\log(\lambda) = -\log(\theta)$.

Solution 3. Integrate the density over the whole sample space $y \in Y$, if $f_\theta(y)$ is a probability density, we can derive that

$$\begin{aligned}
\int_y f_\theta(y)\mathrm{d}y &= \int_y h(y)\exp\Big(\theta\cdot R(y) - \Psi(\theta)\Big)\mathrm{d}y \\
1 &= \int_y h(y)\exp[\theta\cdot R(y)]/[\exp\Psi(\theta)]\mathrm{d}y \\
\exp\Psi(\theta) &= \int_y h(y)\exp\big(\theta\cdot R(y)\big)\mathrm{d}y
\end{aligned}$$

Therefore, $\Psi(\theta) = \log\int h(y)\exp\big(\theta\cdot R(y)\big)\mathrm{d}y$. Taking its derivative w.r.t. θ we can get

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\theta}\Psi(\theta) =& \frac{1}{\int h(y)\exp\big(\theta\cdot R(y)\big)\mathrm{d}y}\cdot\int h(y)\exp\big(\theta\cdot R(y)\big)R(y)\mathrm{d}y \\
=& \frac{1}{\int h(y)\exp\big(\theta\cdot R(y) - \Psi(\theta)\big)\mathrm{d}y}\cdot\int h(y)\exp\big(\theta\cdot R(y) - \Psi(\theta)\big)R(y)\mathrm{d}y \\
=& \int R(y)h(y)\exp\big(\theta\cdot R(y) - \Psi(\theta)\big)\mathrm{d}y = \mathbb{E}_{p(y|\theta)}R(y)
\end{aligned}$$

Solution 4. The solution procedure goes as follows.

- For the marginal prior density of μ: the solution process is simpler when we use the precision $\tau = 1/\sigma^2$, whose prior distribution is naturally the Gamma distribution with same parameters (λ_σ, α). The marginal prior distribution on μ can be obtained by margining out the precision τ:

$$
\begin{aligned}
p(\mu) &= \int_0^\infty p(\mu,\tau)\mathrm{d}\tau \\
&= \int_0^\infty \sqrt{\frac{\tau\lambda_\mu}{2\pi}} \exp\big[-1/2\tau\lambda_\mu(\mu-\xi)^2\big]\frac{\alpha^{\lambda_\sigma}}{\Gamma(\lambda_\sigma)}\tau^{\lambda_\sigma-1}\exp(-\tau\alpha)\mathrm{d}\tau \\
&= \int_0^\infty \sqrt{\frac{\tau\lambda_\mu}{2\pi}} \exp\big[-\tau(1/2\lambda_\mu(\mu-\xi)^2+\alpha)\big]\frac{\alpha^{\lambda_\sigma}}{\Gamma(\lambda_\sigma)}\tau^{\lambda_\sigma-1}\mathrm{d}\tau
\end{aligned}
$$

Introducing the following notation of $z = [\alpha + 1/2\lambda_\mu(\mu-\xi)^2]\tau$, we can further derive that:

$$
\begin{aligned}
p(\mu) =& \sqrt{\frac{\lambda_\mu}{2\pi}} \cdot \frac{\alpha^{\lambda_\sigma}}{\Gamma(\lambda_\sigma)} \cdot \int_0^\infty \tau^{1/2}\exp(-z)\tau^{\lambda_\sigma-1}\mathrm{d}\tau \\
=& \sqrt{\frac{\lambda_\mu}{2\pi}} \cdot \frac{\alpha^{\lambda_\sigma}}{\Gamma(\lambda_\sigma)} \cdot [\alpha+1/2\lambda_\mu(\mu-\xi)^2]^{-(1/2+\lambda_\sigma+1-1)} \int_0^\infty z^{1/2}\exp(-z)z^{\lambda_\sigma-1}\mathrm{d}z \\
=& \sqrt{\frac{\lambda_\mu}{2\pi}} \cdot \frac{\alpha^{\lambda_\sigma}}{\Gamma(\lambda_\sigma)} \cdot [\alpha+1/2\lambda_\mu(\mu-\xi)^2]^{-\lambda_\sigma-1/2}\Gamma(\lambda_\sigma+1/2) \\
=& \frac{\Gamma(2\lambda_\sigma/2+1/2)}{\Gamma(2\lambda_\sigma/2)} \cdot [1+\frac{\lambda_\mu\lambda_\sigma}{\alpha\cdot 2\lambda_\sigma}(\mu-\xi)^2]^{-(2\lambda_\sigma+1)/2} \cdot \sqrt{\frac{\lambda_\mu}{2\pi\alpha}} \\
=& \frac{\Gamma(2\lambda_\sigma/2+1/2)}{\Gamma(2\lambda_\sigma/2)} \cdot [1+\frac{\lambda_\mu\lambda_\sigma}{\alpha\cdot 2\lambda_\sigma}(\mu-\xi)^2]^{-(2\lambda_\sigma+1)/2} \cdot \frac{1}{\sqrt{\pi}}\frac{1}{\sqrt{2\lambda_\sigma(\alpha/\lambda_\mu\lambda_\sigma)}}
\end{aligned}
$$

Hence, it is a Student's t distribution with parameters $(\xi, \alpha/\lambda_\mu\lambda_\sigma, 2\lambda_\sigma)$.

- For the σ^2: with the same manner, but now integrating over μ, we can get:

$$
\begin{aligned}
p(\sigma^2) =& \int_{-\infty}^\infty p(\mu,\sigma^2)\mathrm{d}\mu \\
=& \mathcal{IG}(\sigma^2|\lambda_\sigma,\alpha)\int_{-\infty}^\infty \mathcal{N}(\mu|\xi,\sigma^2/\lambda_\mu)\mathrm{d}\mu \\
=& \mathcal{IG}(\sigma^2|\lambda_\sigma,\alpha)\times 1 = \mathcal{IG}(\sigma^2|\lambda_\sigma,\alpha)
\end{aligned}
$$

Note that we are using $\mathcal{N}(.|a,b)$ and $\mathcal{IG}(.|a,b)$ to denote the normal and inverse gamma densities, respectively. So its marginal prior density remains an inverse-gamma density with parameters (λ_σ,α).

- Posterior parameters: Given an i.i.d. sequence of samples $\mathcal{D} = (x_1,\ldots,x_n)$, the posterior density $p(\mu, \sigma^2|\mathcal{D})$ is proportional to the prior $p(\mu, \sigma^2)$ times the likelihood $p(\mathcal{D}|\mu, \sigma^2)$:

$$\begin{aligned}
p(\mu, \sigma^2|\mathcal{D}) \propto & p(\mu, \sigma^2)p(\mathcal{D}|\mu, \sigma^2) \\
=& \mathcal{N}(\mu|\xi, \sigma^2/\lambda_\mu) \cdot \mathcal{IG}(\sigma^2|\lambda_\sigma, \alpha) \cdot \prod_{i=1}^n \mathcal{N}(x_i|\mu, \sigma^2) \\
\propto & \big[(\sigma^2)^{-1/2} \exp(-\frac{\lambda_\mu}{2\sigma^2}(\mu-\xi)^2)\big] \cdot \big[(\sigma^2)^{-\lambda_\sigma-1} \exp(-\alpha/\sigma^2)\big] \\
& \cdot \big[(\sigma^2)^{-n/2} \exp(-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2)\big] \\
=& (\sigma^2)^{-1/2-n/2-\lambda_\sigma-1} \cdot \exp(-\alpha/\sigma^2) \\
& \cdot \exp(-\frac{\lambda_\mu}{2\sigma^2}(\mu-\xi)^2 - \frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2)
\end{aligned} \tag{1}$$

Let us denote the sample mean $\hat{\mu} = \frac{1}{n}\sum_{i=1}^n x_i$ and variance $\hat{\sigma^2} = \frac{1}{n}\sum_{i=1}^n (x_i - \hat{\mu})^2$, we can express the summation term $\sum_{i=1}^n (x_i-\mu)^2$ as the following:

$$\begin{aligned}
\sum_{i=1}^n (x_i-\mu)^2 =& \sum_{i=1}^n (x_i - \hat{\mu} + \hat{\mu} - \mu)^2 \\
=& n\hat{\sigma^2} + n(\hat{\mu}-\mu)^2 + 2\sum_{i=1}^n (x_i-\hat{\mu})(\hat{\mu}-\mu) \\
=& n\hat{\sigma^2} + n(\hat{\mu}-\mu)^2.
\end{aligned} \tag{2}$$

Plugging Eq (2) back into Eq (1), we can organize and get the following:

$$\begin{aligned}
p(\mu, \sigma^2|\mathcal{D}) \propto & (\sigma^2)^{-1/2-n/2-\lambda_\sigma-1} \cdot \exp(-\alpha/\sigma^2) \\
& \cdot \exp(-\frac{\lambda_\mu}{2\sigma^2}(\mu-\xi)^2 - \frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2) \\
=& (\sigma^2)^{-1/2-n/2-\lambda_\sigma-1} \cdot \exp[-(\alpha + \frac{1}{2}n\hat{\sigma^2})/\sigma^2] \\
& \cdot \exp(-\frac{1}{2\sigma^2}\big[\lambda_\mu(\mu-\xi)^2 + n(\hat{\mu}-\mu)^2\big])
\end{aligned} \tag{3}$$

By completing the square, we can simplify the expression inside the final exponential term as:

$$\begin{aligned}\lambda_\mu(\mu-\xi)^2+n(\hat{\mu}-\mu)^2 =&\lambda_\mu\mu^2-2\lambda_\mu\mu\xi+\lambda_\mu\xi^2+n\hat{\mu}^2-2n\hat{\mu}\mu+n\mu^2\\ =&\mu^2(\lambda_\mu+n)-2\mu(\lambda_\mu\xi+n\hat{\mu})+\lambda_\mu\xi^2+n\mu^2\\ =&(\lambda_\mu+n)(\mu-\frac{\lambda_\mu\xi+n\hat{\mu}}{\lambda_\mu+n})^2+\frac{\lambda_\mu n(\hat{\mu}-\xi)^2}{\lambda_\mu+n}\end{aligned} \tag{4}$$

Plugging Eq (4) back into Eq (3), we can get:

$$\begin{aligned}p(\mu,\sigma^2|\mathcal{D}) \propto&(\sigma^2)^{-1/2-n/2-\lambda_\sigma-1}\cdot\exp[-(\alpha+\frac{1}{2}n\hat{\sigma^2})/\sigma^2]\\ &\cdot\exp(-\frac{1}{2\sigma^2}\left[\lambda_\mu(\mu-\xi)^2+n(\hat{\mu}-\mu)^2\right])\\ =&(\sigma^2)^{-1/2-n/2-\lambda_\sigma-1}\cdot\exp[-(\alpha+\frac{1}{2}n\hat{\sigma^2})/\sigma^2]\cdot\exp(-\frac{1}{2\sigma^2}\frac{\lambda_\mu n(\hat{\mu}-\xi)^2}{\lambda_\mu+n})\\ &\cdot\exp(-\frac{1}{2\sigma^2}\left[(\lambda_\mu+n)(\mu-\frac{\lambda_\mu\xi+n\hat{\mu}}{\lambda_\mu+n})^2\right])\\ =&(\sigma^2)^{-1/2-n/2-\lambda_\sigma-1}\cdot\exp[-\frac{1}{\sigma^2}(\alpha+\frac{1}{2}n\hat{\sigma^2}+\frac{\lambda_\mu n(\hat{\mu}-\xi)^2}{2\lambda_\mu+2n})]\\ &\cdot\exp(-\frac{1}{2\sigma^2}\left[(\lambda_\mu+n)(\mu-\frac{\lambda_\mu\xi+n\hat{\mu}}{\lambda_\mu+n})^2\right])\end{aligned} \tag{5}$$

It is in the form of a normal-inverse-gamma product:

$$p(\mu,\sigma^2|\mathcal{D}) =\mathcal{N}\left(\mu|\frac{\lambda_\mu\xi+n\hat{\mu}}{\lambda_\mu+n},\sigma^2/(\lambda_\mu+n)\right)\cdot\mathcal{IG}\left(\sigma^2|\lambda_\sigma+n/2,\alpha+\frac{1}{2}n\hat{\sigma^2}+\frac{\lambda_\mu n(\hat{\mu}-\xi)^2}{2\lambda_\mu+2n}\right)$$

Solution 5. Note that

$$\begin{aligned}\mathrm{Var}[\bar{X}] =&\mathrm{Var}\left[\frac{1}{N}\sum_{n=1}^N X_n\right]=\frac{1}{N^2}\mathrm{Var}\left[\sum_{n=1}^N X_n\right]\\ =&\frac{1}{N^2}\left(\sum_{n=1}^N\sum_{n'=1}^N\mathrm{Cov}[X_n,X_{n'}]\right)\\ =&\frac{1}{N^2}\left(\sum_{n=1}^N\mathrm{Var}[X_n]\right)+\frac{2}{N^2}\left(\sum_{n=1}^{N-1}\sum_{m=1}^{N-n}\mathrm{Cov}[X_n,X_{n+m}]\right)\end{aligned}$$

For a zero-mean stationary process, $\mathrm{Var}[X_n] = \sigma^2 \quad \forall n$, and $\mathrm{Cov}[X_n, X_{n+m}] = \sigma^2 \rho_m \quad \forall n, m$, we can further derive that:

$$\begin{aligned}
\mathrm{Var}[\bar{X}] =& \frac{1}{N^2}\left(\sum_{n=1}^{N} \mathrm{Var}[X_n]\right) + \frac{2}{N^2}\left(\sum_{n=1}^{N-1}\sum_{m=1}^{N-n} \mathrm{Cov}[X_n, X_{n+m}]\right) \\
=& \frac{1}{N}\sigma^2 + \frac{2}{N^2}\left(\sum_{n=1}^{N-1}\sum_{m=1}^{N-n} \sigma^2 \rho_m\right) \\
=& \frac{1}{N}\sigma^2 + \frac{2}{N^2}\Big((N-1)\sigma^2\rho_1 + (N-2)\sigma^2\rho_2 + \ldots + \sigma^2\rho_{N-1}\Big) \\
=& \frac{1}{N}\sigma^2 + \frac{2}{N^2}\sigma^2 \sum_{m=1}^{N-1}(N-m)\rho_m
\end{aligned}$$

It can be smaller than the one corresponding to uncorrelated X_n's (which is σ^2/N) if negative correlations exist, for example, consider the following process:

$$\rho_m = \begin{cases} 1 & m = 0 \\ -0.1 & m = 1 \\ 0 & m > 1 \end{cases}.$$

Solution 6. The figures should look like Figure 1.

Solution 7. The answer is given as follows.

(a) Note that the trace (tr) is the sum of a square matrix's diagonal components, and the trace trick means $tr(AB) = tr(BA)$ if all dimensions work out, we can derive:

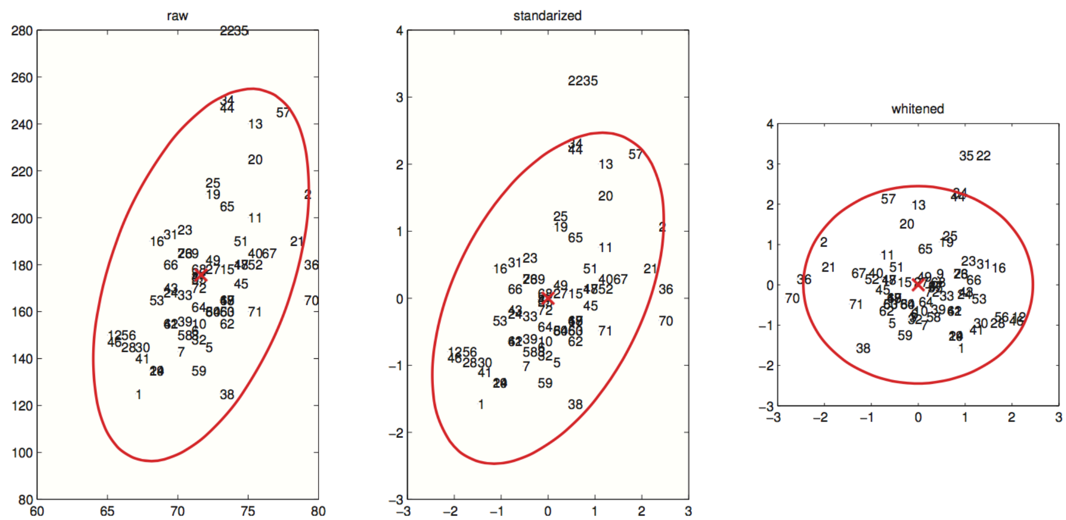

Figure 1: raw data & standardized data & whitened data

$$
\begin{aligned}
\log p(\mathcal{D}|\hat{\Sigma},\hat{\mu}) &= \sum_{i=1}^{N} \log p(\boldsymbol{x}_i|\hat{\Sigma},\hat{\mu}) \\
&= \sum_{i=1}^{N} \Big[-\frac{1}{2}(\boldsymbol{x_i}-\hat{\mu})^T\hat{\Sigma}^{-1}(\boldsymbol{x_i}-\hat{\mu}) - \frac{1}{2}\log(|\det\hat{\Sigma}|)\Big] \\
&= -\frac{1}{2}\sum_{i=1}^{N}(\boldsymbol{x_i}-\hat{\mu})^T\hat{\Sigma}^{-1}(\boldsymbol{x_i}-\hat{\mu}) - \frac{N}{2}\log(|\det\hat{\Sigma}|) \\
&= -\frac{N}{2}\sum_{i=1}^{N}\frac{1}{N}tr\Big((\boldsymbol{x_i}-\hat{\mu})^T\hat{\Sigma}^{-1}(\boldsymbol{x_i}-\hat{\mu})\Big) - \frac{N}{2}\log(|\det\hat{\Sigma}|) \\
&= -\frac{N}{2}tr\Big(\hat{\Sigma}^{-1}\frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{x_i}-\hat{\mu})(\boldsymbol{x_i}-\hat{\mu})^T\Big) - \frac{N}{2}\log(|\det\hat{\Sigma}|) \\
&= -\frac{N}{2}tr\big(\hat{\Sigma}^{-1}\hat{S}\big) - \frac{N}{2}\log(|\det\hat{\Sigma}|)
\end{aligned}
$$

(b) The MLE estimate for the mean and the full covariance matrix are $\hat{\mu}_{ML} = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{x}_i$, and $\hat{\Sigma}_{ML} = \frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{x_i} - \hat{\mu}_{ML})(\boldsymbol{x_i} - \hat{\mu}_{ML})^T$, respectively. Now note that $\hat{\Sigma}_{ML}^{-1}\hat{S} = \boldsymbol{I}$, where $\boldsymbol{I}$ is a D-dimensional identity matrix. For the number of free parameters d, the mean vector contributes D parameters and the covariance matrix contributes $D(D+1)/2$ parameters because of symmetry, hence $d = D + D(D+1)/2$. Then

$$
\begin{aligned}
BIC &= \log p(\mathcal{D}|\hat{\Sigma}_{MLE}, \hat{\mu}_{MLE}) - \frac{d}{2}\log(N) \\
&= -\frac{N}{2}D - \frac{N}{2} log(|\det \hat{\Sigma}_{MLE}|) - \frac{D + D(D+1)/2}{2}\log(N).
\end{aligned}
$$

(c) We can observe that the relationship $\hat{\Sigma}_{ML}^{-1}\hat{S} = \boldsymbol{I}$ holds even for restricted diagonal case. For the number of free parameters d, the mean vector and the covariance matrix both contribute D parameters (only the diagonal elements for the covariance matrix), hence $d = 2D$. Then

$$
\begin{aligned}
BIC &= \log p(\mathcal{D}|\hat{\Sigma}_{MLE}, \hat{\mu}_{MLE}) - \frac{d}{2}\log(N) \\
&= -\frac{N}{2}D - \frac{N}{2} log(|\det \hat{\Sigma}_{MLE}|) - D\log(N).
\end{aligned}
$$